# Prediction and Analysis of the Factors That Influence Volume of Air Passenger at Banyuwangi Airport using Econometric

Suwander Husada[1], Endroyono[1] and Yoyon K. Suprapto[1]

[1]*Department of Electrical Engineering, Institute of Institute Teknologi Sepuluh Nopember, Surabaya, Indonesia*

Keywords: Econometric, Passenger Prediction, Airport Masterplan.

Abstract: The volume of passengers at the airport in the future needs to be predicted. From several air passenger prediction methods, one of them is the econometric method. This method has the excellence of long-term predictions and gaining knowledge of economic factors that affect the volume of air passengers. In this study, several economic factors, such as hotel occupancy, population, rupiah exchange rate, tourism visitor, inflation, and the Banyuwangi Consumer Price Index (CPI), were analyzed. Pre-analysis data needed before building the model, including data selection, data description, data pattern recognition, and data completeness. The best econometric model is obtained from the results of the combination test of each variable. The best econometric model is used to predict the volume of air passengers in Banyuwangi airport for the next 20 years. From the results, three factors significantly effect the volume of air passengers, that is, hotel occupancy affect 0.301%, population affect 0.132%, and the rupiah exchange rate affect -0.481%. To evaluate the accuracy of predictions using the Mean Absolute Percentage Error (MAPE) with a value of 15,208 %.

## 1 INTRODUCTION

Indonesia Civil Avition Law No.1 2009, it is stated that the airport, one of the transportation network nodes. It encourages and supports industrial and trade activities. It encourages and supports industrial and trade activities. The economic stability of Banyuwangi Regency has an impact on the growth of human and goods mobilization from origin to destination through air transportation modes. Referring to the Banyuwangi Airport Master Plan, phase II is estimated to serve aircraft passengers around 272,500 passengers per year, but in 2017 volume of air passengers at Banyuwangi Airport is 188,949 thousand, and by the end of 2018 air passengers volume is 366,155 people. There is a significant difference between the master plan projections and existing passenger. Banyuwangi Airport has the potential to experience density and result in a decrease in the quality of services for users of air transportation services. In this condition, there is a high need to evaluate and projections air volume passengers in the future.

The purpose of this study is to predict the volume of air passengers for the next 20 years and analyze the factors that effect the volume of passengers at Banyuwangi Airport. According result of analysis, it was obtaining knowledge about the factors that influence the volume of air passengers at Banyuwangi Airport. The study result can be used as a reference for policy development.

## 2 LITERATURE REVIEW

Economic factors affect the volume of passengers at certain airports, regions or cities. The most common economic indicators are Gross Domestic Product (GDP), Per capita GDP, Population (Pop), Income, and Per Capita Income (Guo and Zhong, 2017). There are many studies that discuss the relationship between air passenger volume and economic variables.



Husada, S., Endroyono, . and Suprapto, Y.
Prediction and Analysis of the Factors That Influence Volume of Air Passenger at Banyuwangi Airport using Econometric.
DOI: 10.5220/0009880301730179
In *Proceedings of the 2nd International Conference on Applied Science, Engineering and Social Sciences (ICASESS 2019)*, pages 173-179
ISBN: 978-989-758-452-7

Figure 1. Provides insight into several economic indicators that have been used to estimate the volume of air passenger.

| Author | Methods | Variable |
|---|---|---|
| R. GUO, ZW ZHONG, MATTER: in 2017 | Econometric models | of China's GDP exchange rate of SGD to CNY Indian CPI GDP Purchase Rate Community tourism |
| Bilqis Amaliah, Azizha Zeinita, Erma Suryani: 2016 | Dynamics Simulation Econometric models | GDP Population Terminal |
| Vassilios A. Profillidis Population, George N. Botzoris, (2006) | General to Specific Approach, Econometric models, | GDP, index of car ownership, car usage fees, fuel costs of railway trainsLowest competitive Rates Travel time |
| Adetayo Olaniyi Adeniran 2017 ,, Adedayo Ayomide | regression analysis of Ordinary Least Squares | naira currency exchange rate GDP, CPI |

Figure 1: Literature Review.

# 3 METHODOLOGY

To solve the problem, the process is carried out referring to the methodology illustrated in Figure 2.

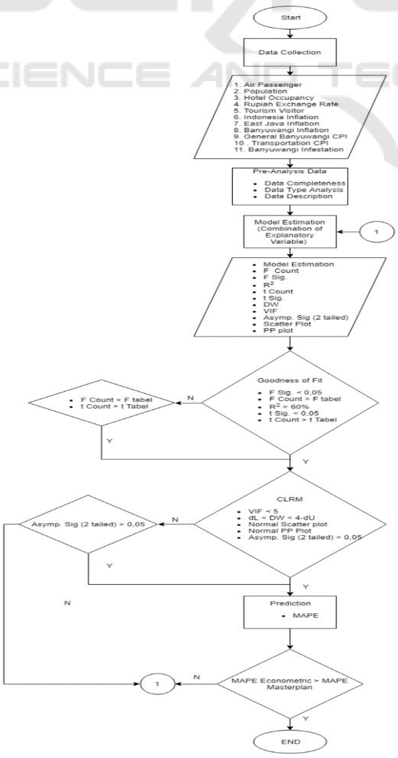



Figure 2: Methodology flow chart

## 3.1 Data Collection

Data in this study was collected from agencies, and website agencies, as shown in Figure 3.

| No | Variable Year | Source | Num. of Sample |
|---|---|---|---|
| **Dependent Variable** | | | |
| 1 | Air Passenger 2012-2017 | Banyuwangi Airport | 72 |
| **Explanatory Variable** | | | |
| 1 | Population 2012-2017 | Dinas Kependudukan dan Catatan Sipil Kabupaten Banyuwangi | 72 |
| 2 | Hotel Occupancy 2012-2017 | Dinas Kebudayaan dan Pariwisata Kabupaten Banyuwangi | 72 |
| 3 | Rupiah exchange rate 2012-2017 | https://www.bi.go.id/ | 72 |
| 4 | Tourism visitor 2013-2017 | Dinas Kebudayaan dan Pariwisata Kabupaten Banyuwangi | 60 |
| 5 | National Inflation 2013-2017 | https: / /www.bi.go.id/ | 60 |
| 6 | East Java Inflation 2013-2017 | https://jatim.bps.go.id/ | 60 |
| 7 | Banyuwangi Inflation 2013-2017 | BPS Banyuwangi | 60 |
| 8 | General Banyuwangi CPI and Transportation CPI 2013-2017 | https://banyuwangikab.bps.go.id/ | 60 |
| 9 | Banyuwangi Infestation 2012-2017 | Dinas penanaman modal dan pelayanan terpadu satu pintu Banyuwangi | 6 |

Figure 3: Dependent Variable and Explanatory Variable.

## 3.2 Pre-analysis Data

Pre-analysis of data consists of :

### 3.2.1 Data Completeness

Completeness of data includes independent variables, explanatory variables, number of data, and year, shown in Figure 3.

### 3.2.2 Data Type Analysis

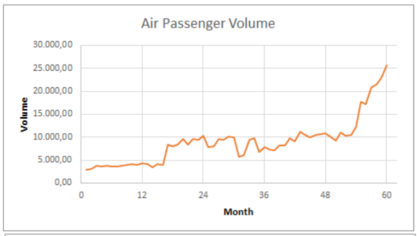
Data type in this research:

- Data Type : Quantitative data
- Data Classified : secondary data
- Measurement : Interval – Ratio

### 3.2.3 Description of Data Graphically

To simplify data analysis, detect trends, seasonal and data patterns, the data is presented using a line chart graph, where the X axis for time (month), and the Yaxis for volume.

Example of graphically shown in Figure 4.



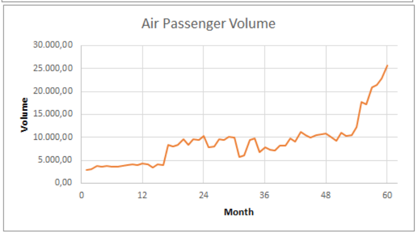




Figure 4: Description of data graphically

There are several data that have the potential to be included in the model, comprise:

- Investment data of Banyuwangi in 2012-2017, the data has annual intervals so that it requires interpolation of monthly intervals. And it is feared that it will cause inaccuracies in the model.
- Banyuwangi Inflation, there are similarities between Banyuwangi inflation data and East Java inflation, In this study used East Java inflation.

#### 3.2.4 Data Description Numerically

Data description numerically is shown in Figure 5. to get the mean, standard deviation, and Number of Sample (N).

| | Mean | Std. Deviation | N |
|---|---|---|---|
| Air Passenger | 7413,0833 | 4797,915 | 72 |
| Hotel Occupancy | 48696,6667 | 7487,596 | 72 |
| Rupiah exchange rate | 12003,5458 | 1649,372 | 72 |
| Population | 1652723,3056 | 25432,931 | 72 |
| Tourism visitor | 202333,1167 | 143462,576 | 60 |
| National Inflation | 5,4715 | 1,926 | 60 |
| East Java Inflation | 0,4112 | 0,592 | 60 |
| General CPI | 117,0058 | 6,384 | 60 |
| Transportation CPI | 115,4777 | 7,850 | 60 |

Figure 5: Descriptive Statistic Table.

### 3.3 Model Estimation

Regression analysis, implemented in equation (1), is needed to measure the strength and direction of passenger relationships and explanatory variables. To determine constants, using the Ordinary Least Squares (OLS) method. In linear regression analysis with OLS, it is necessary to pass the test of the classical linear regression model (CLRM).

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \qquad (1)$$

Y = Dependent Variable $\beta_0$ = Constants $\beta_1, \beta_2 \ldots \beta_n$ = Estimated regression coefficient $x_1$, $x_2$, . . . $x_n$ = Explanatory Variable

Coefficients table is shown in Figure 6 produces values of : $\beta_0, \beta_1, \beta_2 \ldots \beta_n$ , t count, t Sig., Variance Inflation Factor (VIF) value.

| Model | Unstandardized Coefficients | | *t* | *Sig.* | Collinearity Statistics |
|---|---|---|---|---|---|
| | *β* | *Std.Error* | | | *VIF* |
| (Constant) | -220300,605 | 26797,541 | -8,221 | 0,000 | |
| Hotel Occupation | 0,301 | 0,037 | 8,150 | 0,000 | 1,784 |
| Rupiah exchange rate | -0,481 | 0,268 | -1,797 | 0,077 | 4,545 |
| Population | 0,132 | 0,018 | 7,328 | 0,000 | 4,918 |

Figure 6: Coefficients table.

Anova table is shown Figure 7 produces values of : Degree of Freedom (df), F count, and F Sig

| Model | Sum of Square | df | Mean Squre | F | Sig. |
|---|---|---|---|---|---|
| Regression | 1427150815,231 | 3 | 475716938,410 | 156,072 | 0,000 |
| Residual | 207268348,269 | 68 | 3048063,945 | | |
| Total | 1634419163,500 | 71 | | | |

Figure 7: Anova Table.

Model Summary is sown in Figure 8 .produces values of: R , R2 , Durbin Watson (DW)

| Model | R | $R^2$ | Std. Error of the Estimate | df 1 | df 2 | Durbin Watson |
|---|---|---|---|---|---|---|
| 1 | 0,934 | ,873 | 1745,871 | 3 | 68 | 1,611 |

Figure 8: Model Summary Table.

### 3.4 Goodness of Fit / Model Feasibility Test

The accuracy of the regression model obtained in estimating the actual value can be measured by a simultaneous regression coefficient test (F-Test), coefficient of determination (R2), and test individual coefficients test (t-Test).

#### 3.4.1 F-test

The F-test is used to determine whether the Explanatory variable $X_1$, $x_2$,...$x_n$ has a significant influence on the dependent variable Y simultaneously.

- Compare the F Sig to significance level
  - if value of F Sig. $< 0.05$, explanatory variables have a significant effect simultaneously on dependent variable.

- if value of F Sig. $>$ 0.05, explanatory variables have no significant effect simultaneously on dependent variable, or compare F count to t table value.

- Compare t count to t table value
  - if F count $>$ F table value, Explanatory variables have a significant effect simultaneously on dependent variable.
  - if F count $<$ F table value, Explanatory variables have no significant effect simultaneously on dependent variable.

#### 3.4.2 Coefficient of Determination ($R^2$)

Coefficient of Determination using to measure how far the ability of the model explains the dependent variable (Kuncoro, 2001). Implemented in the equation (2).

$$Coefficient of Determination = (R)^2.100\% \quad (2)$$

#### 3.4.3 Individual Coefficients Test (t-Test)

The t-test is used to find out how far the explanatory variables effect the dependent variables individually (Kuncoro, 2001) If t Sig. $<$ Significance level ($\alpha$), the model estimation is declared feasible (Kuncoro, 2001).

- Compare the t Sig to significance level (Siregar, 2013).
  - if value of t Sig. $<$ 0.05, explanatory variables have a significant effect individually on dependent variable.
  - if the value of t Sig. $>$ 0.05, explanatory variables have no significant effect individually on dependent variable. or Compare t count and t table value.
- Compare t count to t Table value (Siregar, 2013).
  - if t count $\geq$ t table value, explanatory variables have a significant effect individually on dependent variable.
  - If -t table value $\leq$ t count $\leq$ t table value explanatory variables have a significant effect individually on dependent variable.

### 3.5 Classical Linear Regression Models (CLRM)

The model must pass the CLRM, comprise:

#### 3.5.1 Multicollinearity

Multicollinearity is a state of high intercorrelations among the explanatory variables. The model is declared free from multicollinearity if the VIF value $<$ 5 . Using VIF value.

- if VIF value $<$ 5, there are no symptoms of Multicollinearity.
- if VIF value $>$ 5, there are symptoms of Multicollinearity.

#### 3.5.2 Autocorrelation / Independence of Observations

In time series data, autocorrelation is defined the error for one time period t is correlated with the error for a subsequent time period t-1. To measure of autocorrelation uses durbin watson test. Compare the DW value with the value of dL and dU on the DW table level. The value of dL and dU are shown in Figure 9.

1. dU $<$ DW$<$ 4-dU item If DW value fall in inconclusive area, uses Run test to measure of autocorrelation (Profillidis and Botzoris, 2006) Figure 10.

| T | K | dL | dU |
|---|---|---|---|
| 72 | 2 | 1.58949 | 1.64571 |
| 72 | 3 | 1.56112 | 1.64571 |
| 72 | 4 | 1.53226 | 1.70539 |
| 72 | 5 | 1.50293 | 1.73664 |

Figure 9: DW Table Level.

T : Number of samples K : Number of variables dL: Lower limit of Durbin Watson dU: Upper limit of Durbin Watson

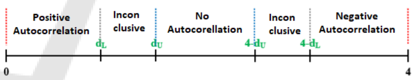


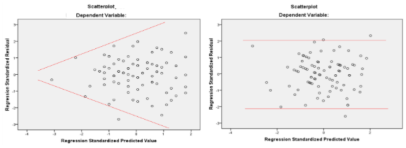

Figure 10: Description of Autocorrelation Using DW

#### 3.5.3 Run Test

Run test is an alternative test to test the autocorrelation if the DW value fall in conculsion area. The test applied by run the randomization test (Profillidis and Botzoris, 2006). Decision in run test, if Asymp. Sig. (2-tailed) $>$ 0,05 , there are no symptoms of autocorrelation.

#### 3.5.4 Heteroscedasticity

The Heteroscedasticity test is a test that assesses whether there is an inequality of variance from the residual for all observations in the linear regression model (Kuncoro, 2001). Heteroscedasticity test analyzes the Scatter plot. The Scatter Plot Graph as seen as Figure 11 created with y and x axis,where x axis is Studentized Residual (SRESID) and y axis is Standardized Predicted Value (ZPRED)".

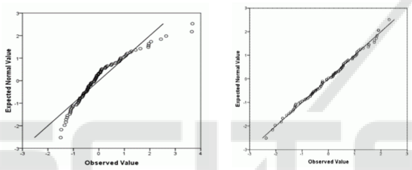



A B

Figure 11: A There are symptoms of heteroskedasticity, B No Symptoms of Heteroskedasticity

#### 3.5.5 Normality

The results of the normality test by observing the distribution of Probability Plots (PP Plots). If the distribution of these points approaches or is in a straight line (diagonal), then it is said that (residual data) is normally distributed, but if the distribution of these points is far from the line, then the skewed distribution as seen as Figure 12.




A B

Figure 12: A PP plot with skewed distribution, B Normal PP Plot

### 3.6 Predictions

Prediction is the process of systematically estimating something that is most likely to occur in the future based on past and present information.

- Winter additive exponential methode Three components winter additive exponential smoothing implemented in equation (3), (4), (5), and (6) (In, 2017):

1. Prediction

$$Y_{t+m} = S_t + mb_t + l_{t-L+m} \tag{3}$$

2. Overall Smoothing

$$S_t = a(X_t - l_{t-1}) + (1-a)(S_t + b_{t-1}) \tag{4}$$

3. Smoothing trend

$$b_t = \beta(S_t - S_{t-1}) + (1-\beta)b_t - 1 \tag{5}$$

4. Seasonal Refinement

$$l_t = Y(X_t - S_t) + (1-Y)l_{t-1} \tag{6}$$

Where :
$S_t$ = Smoothing value for t period.
$X_t$ = Actual value in the final period t
$b_t$ = Value of trend smoothing
$\alpha$ = Smoothing parameters for data ($0 < \alpha < 1$)
Y = Seasonal smoothing parameters ($0 < y < 1$)
$\beta$ = Smoothing parameters for trends ($0 < \beta < 1$)
$l_t$ = Seasonal adjustment factors
L = Seasonal length
$Y_t + m$ = Forcasting for m periode from t
m = Number of periods to forecast

### 3.7 Evaluation of Predictions

The prediction model is validated using Mean Absolute Percentage Error (MAPE) (7).

$$MAPE = \frac{\Sigma \frac{|ei|}{X_i 100}\%}{n} = \frac{\Sigma \frac{|X_i - F_i|}{X_i 100}\%}{n} \tag{7}$$

## 4 RESULTS AND DISCUSSION

From the regression analysis procedure performed, three estimations of the econometric model were built. Each model results from a combination of explanatory variables.

### 4.1 Model 1

Model 1 is built using eight explanatory variables, with 95% Significance level ($\alpha$), and number of samples (N) 60, the details as shown in Figure 13.

Refer to Figure 13, the R2 = 85,9%, meaning the explanatory variable can explain the dependent variables well and has a strong relationship. Explanatory variables have a significant effect simultaneously on dependent variable referring to F.Sig. (0,000) < Significance level (0,05). t Sig of x1, x2, x3 and , x4 < 0,05 it means they have a significant effect individually on dependent variabel. t Sig. of x5, x6, x7 and , x8 > t Significance level, so it need to compare t count to t table. t count of x5, x6, x7 and , x8 < 1,675. It declare , the explanatory variables have no significant effect individually on dependent variable.

According the result of t-test, model 1 not pass feasibility test.

Model 1
$\alpha$ = 95%
N = 60
$Y$ = 1
$x$ = 8
Hotels occupancy($x1$)
Rupiah exchange rate. ($x2$)
Population. ($x3$)
Tourism Visits. ($x4$)
Indonesia Inf ($x5$)
East Java Ifl. ($x6$)
General CPI. ($x7$)
Transport CPI ($x8$)

Regression analysis Results
$Y$= -2.50617,39 +0.315 $x1$ + (-1,411 $x2$) + 0,137 $x3$ + (-0,007 $x4$) + 188,805 $x5$ + (-269,417 $x6$) + 239,97 $x7$ + 558,43 $x8$
$R^2$ = 85,9
F Sig = 0,000
t table = 1.675
F Value = 38.832
F table = 2.13
DW = 1.402
dL = 1.51442
dU = 1.65184

| | t count | t Sig | VIF |
|---|---|---|---|
| $x1$: | 7,662 | 0,000 | 1,610 |
| $x2$ | -2,608 | 0,012 | 8,474 |
| $x3$: | 3,707 | 0,001 | 10,749 |
| $x4$: | -2,574 | 0,013 | 2,400 |
| $x5$: | 0,999 | 0,322 | 2,396 |
| $x6$: | -0,597 | 0,553 | 1,288 |
| $x7$: | 1,108 | 0,273 | 34,553 |
| $x6$: | 0,474 | 0,637 | 15,423 |

Figure 13: Model 1 Information Table

## 4.2 Model 2

Model 2 is Built with four explanatory variables, with 95%Significance level (α), and number of sample (N) 60, the details as shown in Figure 14 .

Model 2
$\alpha$ = 95%
N = 60
$Y$ = 1
$x$ = 4
Hotels occupancy($x1$)
Rupiah exchange rate. ($x2$)
Population. ($x3$)
Tourism Visits. ($x4$)

Regression analysis Results
$Y$= -2.71781,663 +0.324 $x1$ + (-0,526 $x2$) + 0,164 $x3$ + (-0,005 $x4$)
R2 = 84,7
F Sig = 0,000
t table = 1.674
F Value = 75,931
F table = 2.54
DW = 1.402
dL = 1.51442
dU = 1.65184

| | t count | t Sig | VIF |
|---|---|---|---|
| $x1$: | 8,097 | 0,000 | 1,499 |
| $x2$ | -1,718 | 0,091 | 2,688 |
| $x3$: | 7,213 | 0,000 | 4,041 |
| $x4$: | -2,249 | 0,029 | 2,102 |

Figure 14: Model 2 Information Table

Refer to Figure 14, the R2 = 84,7%, meaning the explanatory variable can explain the dependent variables well and has a strong relationship. Explanatory variables have a significant effect simultaneously on dependent variable referring to F.Sig. (0,000) < Significance level (0,05). t Sig of x1, x3and ,x4 < 0,05 it means they have a significant effect individually on dependent, but t Sig of x3> Significance level, so it need to compare t count to t table. t count of x3(7.213) > 1.674. It declare all explanatory variables have significant effect individually on dependent variable. Model 2 pass the Multicollinearity test. According to all VIF value < 5. For autocorrelation test, DW value < dL, it means model 2 have a positive autocorrelation. According the result of autocorrelation test, model 2 not pass CLRM test.

## 4.3 Model 3

Model 3 was Built with three explanatory variables, with 95% Significance level (α), and number of sample (N) 72, the details as shown in Figure 15.

Model 3
$\alpha$ = 95%
N = 72
$Y$ = 1
$x$ = 3
Hotels occupancy($x1$)
Rupiah exchange rate. ($x2$)
Population. ($x3$)

Regression analysis Results
$Y$= -2,27346,580 +0.301 $x1$ + (-0,481 $x2$) + 0,132 $x3$
$R^2$ = 87,3
F Sig = 0,000
t table = 1,670
F Value = 75.931
F table = 2.54
DW = 1.611
dL = 1.53226
dU = 1.70539

| | t count | t Sig | VIF |
|---|---|---|---|
| $x1$: | 8,150 | 0,000 | 1,784 |
| $x2$ | -1,797 | 0,077 | 4,545 |
| $x3$: | 7,328 | 0,000 | 4,918 |

Figure 15: Model 3 Information Table

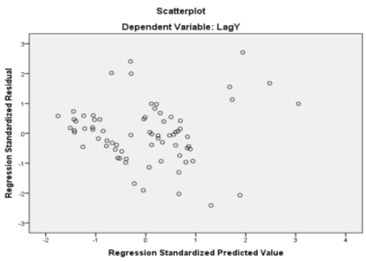
Refer to Figure 8, the R2 = 87,3, meaning the explanatory can explain the dependent variables well and has a strong relationship. Explanatory variables have a significant effect simultaneously on dependent variable referring to F.Sig. (0,000) < Significance level (0,05). t Sig. of x1,and x2 < 0,05 it means they have a significant effect individually on dependent, but t Sig. of x3> t Significance level, so it need to compare t count to t table. t count of x3(-1.797) < - 1.670. It declare all explanatory variables have significant effect individually on dependent variable. Model 3 pass the Multicollinearity test according to all VIF value < 5. For autocorrelation test, DW value < dU, DW value fall in inconclusive area, to solved this condition we uses a Run test. Result of Run test as shown in Figure 16.

Refer to Figure 16, value of Asymp. Sig(2-tailed) 0,476 > 0,05, it means model 3 no autocorrelation. From Figure 12 shows that the point distribution does not form a particular pattern / path, so it can be concluded no heteroscedasticity. Distribution of points from Figure 17 normal P-P Plot.

The above plot is relatively close to a straight line, so it can be concluded that the data residuals are distributed normally.

| Unstandardized | Residual |
|---|---|
| Test Value | 9,55233 |
| Cases < Test Value | 36 |
| Cases ≥ Test Value | 36 |
| Total Cases | 72 |
| Number of Runs | 34 |
| Z | -0,712 |
| Asymp. Sig. (2-tailed) | 0,476 |

Figure 16: Run test result.



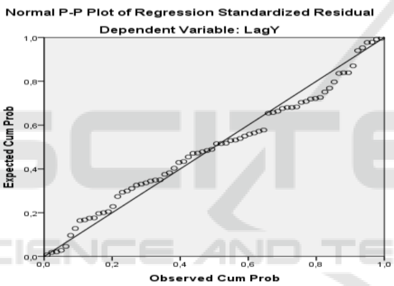




Figure 17: PP Plot

## 4.4 Predictions for the Next 20 Years

After getting the model, we predict each explanatory variable. From the observations of the data on each variable, with the presence of seasonal trends and patterns, the use of predictions suitable for each variable is as follows:

1. Population Winter Additives
2. Hotel Occupancy Winter Additives
3. Currency Exchange Winter Additives

Based on the econometric model 3, the volume of air passengers in Banyuwangi is predicted for the next 20 years. Predictions from 2017 to 2037. Prediction results are shown in Figure 19.

## 4.5 Prediction Validation

Figure 18 shows comparation MAPE value for econometric prediction and Masterplan prediction.

| | MAPE | Explanation |
|---|---|---|
| Masterplan | 22,786 % | The error percentage is 22,768 % |
| Econometric | 15,208 % | The error percentage is 15,208 % |

Figure 18: Comparation MAPE Value.

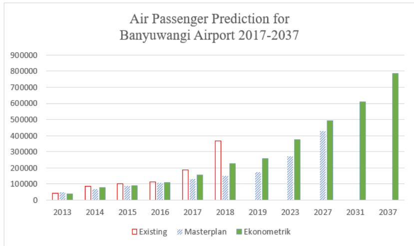



Figure 19: Results of model econometric predictions

# 5 CONCLUSIONS

The results of the analysis; the variables that significantly influence the volume of air passengers at Banyuwangi airport are Hotels Occupancy, Rupiah exchange rates, and Population, with coefficients of value 0, 301, -0.481, and 0.132. A negative sign on the rupiah exchange rate shows an inverse relationship. With the MAPE value of 15,208 %, the prediction of passengers at Banyuwangi Airport using econometrics is declared feasible. Therefore, it is concluded that the economic model provides a good estimate of the volume of air passengers at Banyuwangi Airport.

# REFERENCES

Guo, R. and Zhong, Z. W. (2017). Forecasting air passenger volume in singapore: determining the explanatory variables for econometric models. *MATTER: International Journal of Science and Technology*, 3(1).

In, H. (2017). Pemulusan eksponensial dengan metode holt winter additive damped. *UNHAS Repository*.

Kuncoro, M. (2001). Metodde kuantitatif, teori dan aplikasi untuk bisnis dan ekonomi, upp amp ykpn.

Profillidis, V. A. and Botzoris, G. N. (2006). Econometric models for the forecast of passenger demand in greece. *Journal of Statistics and Management Systems*, 9(1):37–54.

Siregar, S. (2013). Metode penelitian kuantitatif spss. *Yogyakarta: Prenada Media Group*.